Math Camp

Justin Grimmer

Associate Professor
Department of Political Science
Stanford University

September 22nd, 2016

Interpreting Causal Effects

Rubin Causal Model:

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment
- $Y_i(0)$: response under control

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment
- $Y_i(0)$: response under control

Individual Treatment Effect: $Y_i(1) - Y_i(0)$

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment
- $Y_i(0)$: response under control

Individual Treatment Effect: $Y_i(1) - Y_i(0)$
Fundamental problem of causal inference:

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment
- $Y_i(0)$: response under control

Individual Treatment Effect: $Y_i(1) - Y_i(0)$
Fundamental problem of causal inference:
$Y_i(1) - Y_i(0)$

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment
- $Y_i(0)$: response under control

Individual Treatment Effect: $Y_i(1) - Y_i(0)$
Fundamental problem of causal inference:
$Y_i(1)$ -

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment
- $Y_i(0)$: response under control

Individual Treatment Effect: $Y_i(1) - Y_i(0)$

Fundamental problem of causal inference:

- $Y_i(0)$

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment
- $Y_i(0)$: response under control

Individual Treatment Effect: $Y_i(1) - Y_i(0)$
Fundamental problem of causal inference:
$Y_i(1)$ - $Y_i(0)$

Interpreting Causal Effects

Rubin Causal Model:

- $Y_i(1)$: response under treatment
- $Y_i(0)$: response under control

Individual Treatment Effect: $Y_i(1) - Y_i(0)$
Fundamental problem of causal inference:
$Y_i(1)$ - $Y_i(0)$
ATE $= E[Y(1) - Y(0)]$

Observational Studies and Causal Inference

Experimental studies:

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

Because $E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$ and $E[Y(0)] = E[Y(0)|T = 0] = E[Y(0)|T = 1]$

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

Because $E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$ and $E[Y(0)] = E[Y(0)|T = 0] = E[Y(0)|T = 1]$

Observational

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

Because $E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$ and $E[Y(0)] = E[Y(0)|T = 0] = E[Y(0)|T = 1]$

Observational

- Units (people, countries) control their treatment status

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

Because $E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$ and $E[Y(0)] = E[Y(0)|T = 0] = E[Y(0)|T = 1]$

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T=1] - E[Y(0)|T=0]$$

Because $E[Y(1)] = E[Y(1)|T=1] = E[Y(1)|T=0]$ and $E[Y(0)] = E[Y(0)|T=0] = E[Y(0)|T=1]$

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T=1] \neq E[Y(1)|T=0]$, $[EY(0)|T=0] \neq E[Y(0)|T=1]$

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T=1] - E[Y(0)|T=0]$$

Because $E[Y(1)] = E[Y(1)|T=1] = E[Y(1)|T=0]$ and $E[Y(0)] = E[Y(0)|T=0] = E[Y(0)|T=1]$

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T=1] \neq E[Y(1)|T=0]$, $[EY(0)|T=0] \neq E[Y(0)|T=1]$
  - Observables: things we can see, measure, and use in our study

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

Because $E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$ and $E[Y(0)] = E[Y(0)|T = 0] = E[Y(0)|T = 1]$

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T = 1] \neq E[Y(1)|T = 0]$, $[EY(0)|T = 0] \neq E[Y(0)|T = 1]$
  - Observables: things we can see, measure, and use in our study
  - Unobservables: not observables (big problem)

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

Because $E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$ and $E[Y(0)] = E[Y(0)|T = 0] = E[Y(0)|T = 1]$

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T = 1] \neq E[Y(1)|T = 0]$, $[EY(0)|T = 0] \neq E[Y(0)|T = 1]$
  - Observables: things we can see, measure, and use in our study
  - Unobservables: not observables (big problem)
- Naive difference in means will be biased

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

Because $E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$ and $E[Y(0)] = E[Y(0)|T = 0] = E[Y(0)|T = 1]$

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T = 1] \neq E[Y(1)|T = 0]$, $[EY(0)|T = 0] \neq E[Y(0)|T = 1]$
  - Observables: things we can see, measure, and use in our study
  - Unobservables: not observables (big problem)
- Naive difference in means will be biased
- Many, many, potential strategies for limiting bias

Observational Studies and Causal Inference

Experimental studies:

- Treatment under control of analyst
- Random assignment, estimate

$$\widehat{ATE} = E[Y(1)|T = 1] - E[Y(0)|T = 0]$$

Because $E[Y(1)] = E[Y(1)|T = 1] = E[Y(1)|T = 0]$ and $E[Y(0)] = E[Y(0)|T = 0] = E[Y(1)|T = 1]$

Observational

- Units (people, countries) control their treatment status
- Selection: treatment and control groups differ systematically
  - $E[Y(1)|T = 1] \neq E[Y(1)|T = 0]$, $[EY(0)|T = 0] \neq E[Y(0)|T = 1]$
  - Observables: things we can see, measure, and use in our study
  - Unobservables: not observables (big problem)
- Naive difference in means will be biased
- Many, many, potential strategies for limiting bias
- Regression is just one method

The Problem of Selection

$$\text{ATE} = E[Y(1) - Y(0)]$$

The Problem of Selection

$$\begin{aligned}\text{ATE} &= E[Y(1) - Y(0)] \\ &= E[Y(1)] - E[Y(0)]\end{aligned}$$

The Problem of Selection

$$\begin{aligned}
\text{ATE} &= E[Y(1) - Y(0)] \\
&= E[Y(1)] - E[Y(0)] \\
&= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\
&\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\}
\end{aligned}$$

The Problem of Selection

$$\begin{aligned}
\text{ATE} &= E[Y(1) - Y(0)] \\
&= E[Y(1)] - E[Y(0)] \\
&= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\
&\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\}
\end{aligned}$$

There is no assumption free method for estimating quantities in blue

The Problem of Selection

$$
\begin{aligned}
\text{ATE} &= E[Y(1) - Y(0)] \\
&= E[Y(1)] - E[Y(0)] \\
&= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\
&\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\}
\end{aligned}
$$

There is no assumption free method for estimating quantities in blue

Experiments:

The Problem of Selection

$$\begin{aligned}
\text{ATE} &= E[Y(1) - Y(0)] \\
&= E[Y(1)] - E[Y(0)] \\
&= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\
&\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\}
\end{aligned}$$

There is no assumption free method for estimating quantities in blue

Experiments:

- Control assignment, learn about counterfactual values

The Problem of Selection

$$\begin{aligned}
\text{ATE} &= E[Y(1) - Y(0)] \\
&= E[Y(1)] - E[Y(0)] \\
&= \{E[Y(1)|T=1]Pr(T=1) + E[Y(1)|T=0]Pr(T=0)\} \\
&\quad -\{E[Y(0)|T=0]Pr(T=0) + E[Y(0)|T=1]Pr(T=1)\}
\end{aligned}$$

There is no assumption free method for estimating quantities in blue

Experiments:

- Control assignment, learn about counterfactual values

Observational studies:

The Problem of Selection

$$
\begin{aligned}
\text{ATE} &= E[Y(1) - Y(0)] \\
&= E[Y(1)] - E[Y(0)] \\
&= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\
&\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\}
\end{aligned}
$$

There is no assumption free method for estimating quantities in blue

Experiments:

- Control assignment, learn about counterfactual values

Observational studies:

- Assignment not controlled

The Problem of Selection

$$
\begin{aligned}
\text{ATE} &= E[Y(1) - Y(0)] \\
&= E[Y(1)] - E[Y(0)] \\
&= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\
&\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\}
\end{aligned}
$$

There is no assumption free method for estimating quantities in blue

Experiments:

- Control assignment, learn about counterfactual values

Observational studies:

- Assignment not controlled
- Problem: how do we learn about counterfactuals in the face of selection?

Manski Bound

What can we learn in the face of selection?

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

- Smallest ATE: -1

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

- Smallest ATE: -1
- Largest ATE: 1

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

- Smallest ATE: -1
- Largest ATE: 1
- Length: 2

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

- Smallest ATE: -1
- Largest ATE: 1
- Length: 2
- $-1 < ATE < 1$

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

- Smallest ATE: -1
- Largest ATE: 1
- Length: 2
- $-1 < ATE < 1$

Manksi Bounds:

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

- Smallest ATE: -1
- Largest ATE: 1
- Length: 2
- $-1 < ATE < 1$

Manksi Bounds:

- Interval that contains true ATE with probability 1

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

- Smallest ATE: -1
- Largest ATE: 1
- Length: 2
- $-1 < ATE < 1$

Manksi Bounds:

- Interval that contains true ATE with probability 1
- Reduces interval length to 1

Manski Bound

What can we learn in the face of selection?
Suppose that we have a dichotomous dependent variable

- Smallest ATE: -1
- Largest ATE: 1
- Length: 2
- $-1 < ATE < 1$

Manksi Bounds:

- Interval that contains true ATE with probability 1
- Reduces interval length to 1
- Requires no additional assumptions

Manski Bound

Note that:
$0 \leq E[Y(1)|T = 0] \leq 1$

Manski Bound

Note that:
$0 \leq E[Y(1)|T = 0] \leq 1$
And that:
$0 \leq E[Y(0)|T = 1] \leq 1$

Manski Bound

Note that:

$0 \leq E[Y(1)|T = 0] \leq 1$

And that:

$0 \leq E[Y(0)|T = 1] \leq 1$

And recall that,

Manski Bound

Note that:

$0 \leq E[Y(1)|T = 0] \leq 1$

And that:

$0 \leq E[Y(0)|T = 1] \leq 1$

And recall that,

$$\begin{aligned} ATE &= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\ &\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\} \\ &= E[Y(1)|T = 1]\pi + E[Y(1)|T = 0](1 - \pi) \\ &\quad -E[Y(0)|T = 0](1 - \pi) - E[Y(0)|T = 1]\pi \end{aligned}$$

Manski Bound

Note that:

$0 \leq E[Y(1)|T = 0] \leq 1$

And that:

$0 \leq E[Y(0)|T = 1] \leq 1$

And recall that,

$$\begin{aligned} ATE &= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\ &\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\} \\ &= E[Y(1)|T = 1]\pi + E[Y(1)|T = 0](1 - \pi) \\ &\quad -E[Y(0)|T = 0](1 - \pi) - E[Y(0)|T = 1]\pi \end{aligned}$$

So, we can form bounds

Manski Bound

Note that:

$0 \leq E[Y(1)|T = 0] \leq 1$

And that:

$0 \leq E[Y(0)|T = 1] \leq 1$

And recall that,

$$\begin{aligned} ATE &= \{E[Y(1)|T=1]Pr(T=1) + E[Y(1)|T=0]Pr(T=0)\} \\ &\quad -\{E[Y(0)|T=0]Pr(T=0) + E[Y(0)|T=1]Pr(T=1)\} \\ &= E[Y(1)|T=1]\pi + E[Y(1)|T=0](1-\pi) \\ &\quad -E[Y(0)|T=0](1-\pi) - E[Y(0)|T=1]\pi \end{aligned}$$

So, we can form bounds

$$\begin{aligned} E[Y(1)|T=1]\pi - (1-\pi)E[Y(0)|T=0] - \pi \\ < ATE \end{aligned}$$

Manski Bound

Note that:

$0 \leq E[Y(1)|T = 0] \leq 1$

And that:

$0 \leq E[Y(0)|T = 1] \leq 1$

And recall that,

$$
\begin{aligned}
ATE &= \{E[Y(1)|T = 1]Pr(T = 1) + E[Y(1)|T = 0]Pr(T = 0)\} \\
&\quad -\{E[Y(0)|T = 0]Pr(T = 0) + E[Y(0)|T = 1]Pr(T = 1)\} \\
&= E[Y(1)|T = 1]\pi + E[Y(1)|T = 0](1 - \pi) \\
&\quad -E[Y(0)|T = 0](1 - \pi) - E[Y(0)|T = 1]\pi
\end{aligned}
$$

So, we can form bounds

$$
\begin{aligned}
E[Y(1)|T = 1]\pi - (1 - \pi)E[Y(0)|T = 0] - \pi \\
< ATE < \\
E[Y(1)|T =]\pi - (1 - \pi)E[Y(0)|T = 0] + 1 - \pi
\end{aligned}
$$

Manski Bound

The Manski bound

Manski Bound

The Manski bound

$$[\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} - \pi, \quad \text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} + 1 - \pi]$$

and has length 1.

Manski Bound

The Manski bound

$$[\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} - \pi, \\ \text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} + 1 - \pi]$$

and has length 1.

From Keele 2009: Presidential Visits + Congressional Wins

Manski Bound

The Manski bound

$$[\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} - \pi,$$
$$\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} + 1 - \pi]$$

and has length 1.

From Keele 2009: Presidential Visits + Congressional Wins

- $\pi = 0.06$

Manski Bound

The Manski bound

$$[\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} - \pi,\\ \text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} + 1 - \pi]$$

and has length 1.

From Keele 2009: Presidential Visits + Congressional Wins

- $\pi = 0.06$
- $E[Y(1)|T=1] = 18/21 = 0.857$

Manski Bound

The Manski bound

$$[\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} - \pi,$$
$$\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} + 1 - \pi]$$

and has length 1.

From Keele 2009: Presidential Visits + Congressional Wins

- $\pi = 0.06$
- $E[Y(1)|T=1] = 18/21 = 0.857$
- $E[Y(0)|T=0] = 163/(327) = 0.498$

Manski Bound

The Manski bound

$$[\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} - \pi,$$
$$\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} + 1 - \pi]$$

and has length 1.

From Keele 2009: Presidential Visits + Congressional Wins

- $\pi = 0.06$
- $E[Y(1)|T = 1] = 18/21 = 0.857$
- $E[Y(0)|T = 0] = 163/(327) = 0.498$

Bound is then:

Manski Bound

The Manski bound

$$[\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} - \pi,$$
$$\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} + 1 - \pi]$$

and has length 1.

From Keele 2009: Presidential Visits + Congressional Wins

- $\pi = 0.06$
- $E[Y(1)|T = 1] = 18/21 = 0.857$
- $E[Y(0)|T = 0] = 163/(327) = 0.498$

Bound is then:

$$[0.86 \times 0.06 - 0.94 \times 0.498 - 0.06, 0.86 \times 0.06 - 0.94 \times 0.498 + 0.94]$$
$$[-0.476, 0.523]$$

Manski Bound

The Manski bound

$$[\text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} - \pi, \\ \text{Average(Treat)}\pi - (1-\pi)\text{Average(Control)} + 1 - \pi]$$

and has length 1.

From Keele 2009: Presidential Visits + Congressional Wins

- $\pi = 0.06$
- $E[Y(1)|T=1] = 18/21 = 0.857$
- $E[Y(0)|T=0] = 163/(327) = 0.498$

Bound is then:

$$[0.86 \times 0.06 - 0.94 \times 0.498 - 0.06, 0.86 \times 0.06 - 0.94 \times 0.498 + 0.94]$$
$$[-0.476, 0.523]$$

- Further assumptions can narrow bounds

Manski Bound

The Manski bound

$$[\text{Average}(\text{Treat})\pi - (1-\pi)\text{Average}(\text{Control}) - \pi, \quad \text{Average}(\text{Treat})\pi - (1-\pi)\text{Average}(\text{Control}) + 1 - \pi]$$

and has length 1.

From Keele 2009: Presidential Visits + Congressional Wins

- $\pi = 0.06$
- $E[Y(1)|T = 1] = 18/21 = 0.857$
- $E[Y(0)|T = 0] = 163/(327) = 0.498$

Bound is then:

$$[0.86 \times 0.06 - 0.94 \times 0.498 - 0.06, 0.86 \times 0.06 - 0.94 \times 0.498 + 0.94]$$

$$[-0.476, 0.523]$$

- Further assumptions can narrow bounds
- Law of Decreasing Credibility: the credibility of inference decreases with the strength of the assumptions maintained

Causal Inference Via Stratification

Selection on observables:

Causal Inference Via Stratification

Selection on observables:

- Suppose that treatment is systematically related to outcomes (not ignorable)

Causal Inference Via Stratification

Selection on observables:

- Suppose that treatment is systematically related to outcomes (not ignorable)
    - $E[Y(1)|D = 1] \neq E[Y(1)|D = 0]$

Causal Inference Via Stratification

Selection on observables:

- Suppose that treatment is systematically related to outcomes (not ignorable)
    - $E[Y(1)|D=1] \neq E[Y(1)|D=0]$
    - $E[Y(0)|D=1] \neq E[Y(0)|D=0]$

Causal Inference Via Stratification

Selection on observables:

- Suppose that treatment is systematically related to outcomes (not ignorable)
  - $E[Y(1)|D = 1] \neq E[Y(1)|D = 0]$
  - $E[Y(0)|D = 1] \neq E[Y(0)|D = 0]$
  - Implies that Naive difference is not an unbiased estimate of ATE.

Causal Inference Via Stratification

Selection on observables:

- Suppose that treatment is systematically related to outcomes (not ignorable)
    - $E[Y(1)|D = 1] \neq E[Y(1)|D = 0]$
    - $E[Y(0)|D = 1] \neq E[Y(0)|D = 0]$
    - Implies that Naive difference is not an unbiased estimate of ATE.
- Suppose there exists dichotomous variable S such that,

Causal Inference Via Stratification

Selection on observables:

- Suppose that treatment is systematically related to outcomes (not ignorable)
  - $E[Y(1)|D = 1] \neq E[Y(1)|D = 0]$
  - $E[Y(0)|D = 1] \neq E[Y(0)|D = 0]$
  - Implies that Naive difference is not an unbiased estimate of ATE.
- Suppose there exists dichotomous variable S such that,
  1) $E[Y(1)|D = 1, S] = E[Y(1)|D = 0, S]$

Causal Inference Via Stratification

Selection on observables:

- Suppose that treatment is systematically related to outcomes (not ignorable)
  - $E[Y(1)|D=1] \neq E[Y(1)|D=0]$
  - $E[Y(0)|D=1] \neq E[Y(0)|D=0]$
  - Implies that Naive difference is not an unbiased estimate of ATE.
- Suppose there exists dichotomous variable S such that,
  1) $E[Y(1)|D=1,S] = E[Y(1)|D=0,S]$
  2) $E[Y(0)|D=1,S] = E[Y(0)|D=0,S]$

Causal Inference Via Stratification

Selection on observables:

- Suppose that treatment is systematically related to outcomes (not ignorable)
    - $E[Y(1)|D=1] \neq E[Y(1)|D=0]$
    - $E[Y(0)|D=1] \neq E[Y(0)|D=0]$
    - Implies that Naive difference is not an unbiased estimate of ATE.
- Suppose there exists dichotomous variable S such that,
    1) $E[Y(1)|D=1,S] = E[Y(1)|D=0,S]$
    2) $E[Y(0)|D=1,S] = E[Y(0)|D=0,S]$

Proposition

Suppose there exists S such that 1) and 2) hold. Then, we can obtain unbiased estimates for

1) $E[ATE|S=1] \equiv$ *ATE, Given* $S=1$
2) $E[ATE|S=0] \equiv$ *ATE, Given* $S=0$
3) *ATE*

Idea:

Idea:

1) Average Treatment effect with strata, $S = 1$

Idea:

1) Average Treatment effect with strata, $S = 1$
   - Average(Treatment) - Average(Control) , for all units with $S = 1$

Idea:

1) Average Treatment effect with strata, $S = 1$
   - Average(Treatment) - Average(Control) , for all units with $S = 1$
2) Average Treatment effect with strata, $S = 0$

Idea:

1) Average Treatment effect with strata, $S = 1$
   - Average(Treatment) - Average(Control) , for all units with $S = 1$
2) Average Treatment effect with strata, $S = 0$
   - Average(Treatment) - Average(Control) , for all units with $S = 0$

Idea:

1) Average Treatment effect with strata, $S = 1$
   - Average(Treatment) - Average(Control) , for all units with $S = 1$
2) Average Treatment effect with strata, $S = 0$
   - Average(Treatment) - Average(Control) , for all units with $S = 0$
3) Average Treatment Effect

Idea:

1) Average Treatment effect with strata, $S = 1$
   - Average(Treatment) - Average(Control) , for all units with $S = 1$
2) Average Treatment effect with strata, $S = 0$
   - Average(Treatment) - Average(Control) , for all units with $S = 0$
3) Average Treatment Effect
   - Calculate Average(Treatment) - Average(Control) within each strata, $E[ATE|S]$

Idea:

1) Average Treatment effect with strata, $S = 1$
   - Average(Treatment) - Average(Control) , for all units with $S = 1$
2) Average Treatment effect with strata, $S = 0$
   - Average(Treatment) - Average(Control) , for all units with $S = 0$
3) Average Treatment Effect
   - Calculate Average(Treatment) - Average(Control) within each strata, $E[ATE|S]$
   - $ATE = \sum_{s=0}^{1} E[ATE|S] \times Pr(S)$

Proof

$$E[Y(1)|S = 1, D = 1] - E[Y(0)|S = 0, D = 0]$$

Proof

$$E[Y(1)|S=1,D=1] - E[Y(0)|S=0,D=0]$$

$$\begin{aligned} = \quad & \{\Pr(D=1|S=1)E[Y(1)|S=1,D=1] \\ & +\Pr(D=0|S=1)E[Y(1)|S=1,D=1]\} \\ & -\{\Pr(D=1|S=1)E[Y(0)|S=1,D=0] \\ & +\Pr(D=0|S=1)E[Y(0)|S=1,D=0]\} \end{aligned}$$

Proof

$$E[Y(1)|S = 1, D = 1] - E[Y(0)|S = 0, D = 0]$$

$$\begin{aligned}
= & \{\Pr(D = 1|S = 1)E[Y(1)|S = 1, D = 1] \\
& +\Pr(D = 0|S = 1)E[Y(1)|S = 1, D = 1]\} \\
& -\{\Pr(D = 1|S = 1)E[Y(0)|S = 1, D = 0] \\
& +\Pr(D = 0|S = 1)E[Y(0)|S = 1, D = 0]\} \\
= & \{\Pr(D = 1|S = 1)E[Y(1)|S = 1, D = 1] \\
& +\Pr(D = 0|S = 1)E[Y(1)|S = 1, D = 0]\} \\
& -\{\Pr(D = 1|S = 1)E[Y(0)|S = 1, D = 1] \\
& +\Pr(D = 0|S = 1)E[Y(0)|S = 1, D = 0]\}
\end{aligned}$$

Proof

$$E[Y(1)|S=1,D=1] - E[Y(0)|S=0,D=0]$$

$$\begin{aligned}
&= \{\Pr(D=1|S=1)E[Y(1)|S=1,D=1] \\
&\quad +\Pr(D=0|S=1)E[Y(1)|S=1,D=1]\} \\
&\quad -\{\Pr(D=1|S=1)E[Y(0)|S=1,D=0] \\
&\quad +\Pr(D=0|S=1)E[Y(0)|S=1,D=0]\} \\
&= \{\Pr(D=1|S=1)E[Y(1)|S=1,D=1] \\
&\quad +\Pr(D=0|S=1)E[Y(1)|S=1,D=0]\} \\
&\quad -\{\Pr(D=1|S=1)E[Y(0)|S=1,D=1] \\
&\quad +\Pr(D=0|S=1)E[Y(0)|S=1,D=0]\} \\
&= E[ATE|S=1]
\end{aligned}$$

Causal Inference via Stratification: Example

We are interested in the causal effect of incumbency on reelection.

- $T = 1$, Incumbent
- $T = 0$, Challenger

$Y_i(T)$ result of election.
Incumbency obviously not assigned at random.
But suppose we have a dichotomous measure of candidate quality

- $S = 1$, High quality
- $S = 0$, Low quality

And that incumbency is as good as random, given S.
We're interested in obtaining

$$ATE = E[Y(1) - Y(0)]$$

What if we don't condition on candidate quality?

Joint Distribution of Treat and Strata

	$T = 1$	$T = 0$	$P(S)$
$S = 1$	0.4	0.2	0.6
$S = 0$	0.1	0.3	0.4
$P(T)$	0.5	0.5	

Joint Distribution of Treat and Strata

	$T = 1$	$T = 0$	$P(S)$
$S = 1$	0.4	0.2	0.6
$S = 0$	0.1	0.3	0.4
$P(T)$	0.5	0.5	

Potential Outcomes

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\|S$
S = 1	$E[Y(0)\|S = 1] = 0.5$	$E[Y(1)\|S = 1] = 0.7$	0.2
S = 0	$E[Y(0)\|S = 0] = 0.38$	$E[Y(1)\|S = 0] = 0.4$	0.02

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\vert S$
S = 1	$E[Y(0)\vert S = 1] = 0.5$	$E[Y(1)\vert S = 1] = 0.7$	0.2
S = 0	$E[Y(0)\vert S = 0] = 0.38$	$E[Y(1)\vert S = 0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\mid S$
S = 1	$E[Y(0)\mid S=1] = 0.5$	$E[Y(1)\mid S=1] = 0.7$	0.2
S = 0	$E[Y(0)\mid S=0] = 0.38$	$E[Y(1)\mid S=0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Naive difference in means:

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\|S$
S = 1	$E[Y(0)\|S=1] = 0.5$	$E[Y(1)\|S=1] = 0.7$	0.2
S = 0	$E[Y(0)\|S=0] = 0.38$	$E[Y(1)\|S=0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Naive difference in means:

$$E[Y(1)|T=1] = 0.7 \times P(S=1|D=1) + 0.4 \times P(S=0|D=1)$$

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\|S$
S = 1	$E[Y(0)\|S=1] = 0.5$	$E[Y(1)\|S=1] = 0.7$	0.2
S = 0	$E[Y(0)\|S=0] = 0.38$	$E[Y(1)\|S=0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Naive difference in means:

$$\begin{aligned} E[Y(1)|T=1] &= 0.7 \times P(S=1|D=1) \\ &\quad +0.4 \times P(S=0|D=1) \\ &= 0.7 \times \frac{4}{5} + 0.4 \times \frac{1}{5} = 0.64 \end{aligned}$$

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\|S$
S = 1	$E[Y(0)\|S=1] = 0.5$	$E[Y(1)\|S=1] = 0.7$	0.2
S = 0	$E[Y(0)\|S=0] = 0.38$	$E[Y(1)\|S=0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Naive difference in means:

$$
\begin{aligned}
E[Y(1)|T=1] &= 0.7 \times P(S=1|D=1) \\
&\quad + 0.4 \times P(S=0|D=1) \\
&= 0.7 \times \frac{4}{5} + 0.4 \times \frac{1}{5} = 0.64 \\
E[Y(0)|T=0] &= 0.5 \times \frac{2}{5} + 0.3 \times \frac{3}{5} = 0.38
\end{aligned}
$$

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\|S$
S = 1	$E[Y(0)\|S=1] = 0.5$	$E[Y(1)\|S=1] = 0.7$	0.2
S = 0	$E[Y(0)\|S=0] = 0.38$	$E[Y(1)\|S=0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Naive difference in means:

$$
\begin{aligned}
E[Y(1)|T=1] &= 0.7 \times P(S=1|D=1) \\
&\quad +0.4 \times P(S=0|D=1) \\
&= 0.7 \times \frac{4}{5} + 0.4 \times \frac{1}{5} = 0.64 \\
E[Y(0)|T=0] &= 0.5 \times \frac{2}{5} + 0.3 \times \frac{3}{5} = 0.38 \\
E[Y(1)|T=1] - E[Y(0)|T=0] &= 0.64 - 0.38
\end{aligned}
$$

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\|S$
S = 1	$E[Y(0)\|S=1] = 0.5$	$E[Y(1)\|S=1] = 0.7$	0.2
S = 0	$E[Y(0)\|S=0] = 0.38$	$E[Y(1)\|S=0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Naive difference in means:

$$\begin{aligned}
E[Y(1)|T=1] &= 0.7 \times P(S=1|D=1) \\
&\quad + 0.4 \times P(S=0|D=1) \\
&= 0.7 \times \frac{4}{5} + 0.4 \times \frac{1}{5} = 0.64 \\
E[Y(0)|T=0] &= 0.5 \times \frac{2}{5} + 0.3 \times \frac{3}{5} = 0.38 \\
E[Y(1)|T=1] - E[Y(0)|T=0] &= 0.64 - 0.38 \\
&= 0.26
\end{aligned}$$

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	$\text{ATE}\|S$
S = 1	$E[Y(0)\|S=1] = 0.5$	$E[Y(1)\|S=1] = 0.7$	0.2
S = 0	$E[Y(0)\|S=0] = 0.38$	$E[Y(1)\|S=0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Naive difference in means:

$$
\begin{aligned}
E[Y(1)|T=1] &= 0.7 \times P(S=1|D=1) \\
&\quad +0.4 \times P(S=0|D=1) \\
&= 0.7 \times \frac{4}{5} + 0.4 \times \frac{1}{5} = 0.64 \\
E[Y(0)|T=0] &= 0.5 \times \frac{2}{5} + 0.3 \times \frac{3}{5} = 0.38 \\
E[Y(1)|T=1] - E[Y(0)|T=0] &= 0.64 - 0.38 \\
&= 0.26
\end{aligned}
$$

Why?: those with bigger treatment effects more likely to select into treatment

Joint Distribution of Treat and Strata

	T = 1	T = 0	P(S)
S = 1	0.4	0.2	0.6
S = 0	0.1	0.3	0.4
P(T)	0.5	0.5	

Potential Outcomes

	Control	Treat	ATE$\|S$
S = 1	$E[Y(0)\|S=1] = 0.5$	$E[Y(1)\|S=1] = 0.7$	0.2
S = 0	$E[Y(0)\|S=0] = 0.38$	$E[Y(1)\|S=0] = 0.4$	0.02

True $ATE = 0.2 \times 0.6 + 0.02 \times 0.4 = 0.128$

Naive difference in means:

$$
\begin{aligned}
E[Y(1)|T=1] &= 0.7 \times P(S=1|D=1) \\
&\quad + 0.4 \times P(S=0|D=1) \\
&= 0.7 \times \frac{4}{5} + 0.4 \times \frac{1}{5} = 0.64 \\
E[Y(0)|T=0] &= 0.5 \times \frac{2}{5} + 0.3 \times \frac{3}{5} = 0.38 \\
E[Y(1)|T=1] - E[Y(0)|T=0] &= 0.64 - 0.38 \\
&= 0.26
\end{aligned}
$$

Why?: those with bigger treatment effects more likely to select into treatment

Confound effect of T with differences across S

With Stratification

With Stratification

$$
\begin{aligned}
E[Y(1)|T = 1, S = 1] &= 0.7 \\
E[Y(0)|T = 0, S = 1] &= 0.5 \\
E[Y(1)|T = 1, S = 0] &= 0.4 \\
E[Y(0)|T = 0, S = 0] &= 0.38
\end{aligned}
$$

With Stratification

$$\begin{aligned} E[Y(1)|T=1,S=1] &= 0.7 \\ E[Y(0)|T=0,S=1] &= 0.5 \\ E[Y(1)|T=1,S=0] &= 0.4 \\ E[Y(0)|T=0,S=0] &= 0.38 \end{aligned}$$

$\sum_{i=1}^{2} \left(E[Y(1)|S=i,T=1] - E[Y(0)|S=i,T=0]\right) \Pr(S=i) =$

With Stratification

$$
\begin{aligned}
E[Y(1)|T = 1, S = 1] &= 0.7 \\
E[Y(0)|T = 0, S = 1] &= 0.5 \\
E[Y(1)|T = 1, S = 0] &= 0.4 \\
E[Y(0)|T = 0, S = 0] &= 0.38
\end{aligned}
$$

$\sum_{i=1}^{2} \left(E[Y(1)|S = i, T = 1] - E[Y(0)|S = i, T = 0]\right) \Pr(S = i) =$
$(0.7 - 0.5) \times 0.6 + (0.4 - 0.38) \times 0.4 = 0.128$

Stratification and Matching

Exact stratification is Exact Matching

Stratification and Matching

Exact stratification is Exact Matching

Exact Matching, Basic idea:

Stratification and Matching

Exact stratification is Exact Matching

Exact Matching, Basic idea:

- Identify all characteristics (covariates) that affect both outcome and treatment assignment (hint, more than 1!)

Stratification and Matching

Exact stratification is Exact Matching

Exact Matching, Basic idea:

- Identify all characteristics (covariates) that affect both outcome and treatment assignment (hint, more than 1!)
- For all treated units, identify control unit with same characteristics

Stratification and Matching

Exact stratification is Exact Matching

Exact Matching, Basic idea:

- Identify all characteristics (covariates) that affect both outcome and treatment assignment (hint, more than 1!)
- For all treated units, identify control unit with same characteristics
- Exact match: units in the same strata

What do we do?

Most strata are empty, or only a few observations

- Bias-Variance tradeoff
  - Bias: assume same casual effect across strata
  - Variance: assume different causal effect across strata
- Modeling $E[Y|X_1, X_2, X_3, \ldots, X_K]$
  - Nonparametric (loess): different curse of dimensionality problem
  - High dimensional space is sparse, hard to borrow across

Solution: specify a model of how covariates relate to treatment

Linear Regression

Consider one continuous covariate S_i and a continuous dependent variable Y_i.

Linear Regression

Consider one continuous covariate S_i and a continuous dependent variable Y_i.
Assume the following, relationship,

Linear Regression

Consider one continuous covariate S_i and a continuous dependent variable Y_i.
Assume the following, relationship,

$$E[Y_i|S_i] = \beta_0 + \beta_1 S_i$$

Linear Regression

Consider one continuous covariate S_i and a continuous dependent variable Y_i.
Assume the following, relationship,

$$E[Y_i|S_i] = \beta_0 + \beta_1 S_i$$

What does this say?

Linear Regression

Consider one continuous covariate S_i and a continuous dependent variable Y_i.
Assume the following, relationship,

$$E[Y_i|S_i] = \beta_0 + \beta_1 S_i$$

What does this say?

- Stratifying (conditioning): examining means of Y given values of S

Linear Regression

Consider one continuous covariate S_i and a continuous dependent variable Y_i.
Assume the following, relationship,

$$E[Y_i|S_i] \quad = \quad \beta_0 + \beta_1 S_i$$

What does this say?

- Stratifying (conditioning): examining means of Y given values of S
- Borrowing information across bins:

Linear Regression

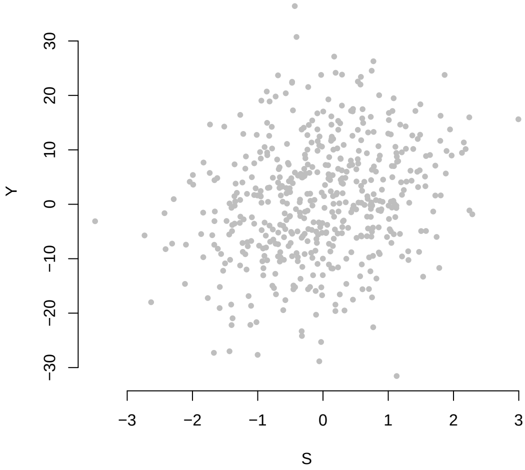
Consider one continuous covariate S_i and a continuous dependent variable Y_i.
Assume the following, relationship,

$$E[Y_i|S_i] = \beta_0 + \beta_1 S_i$$

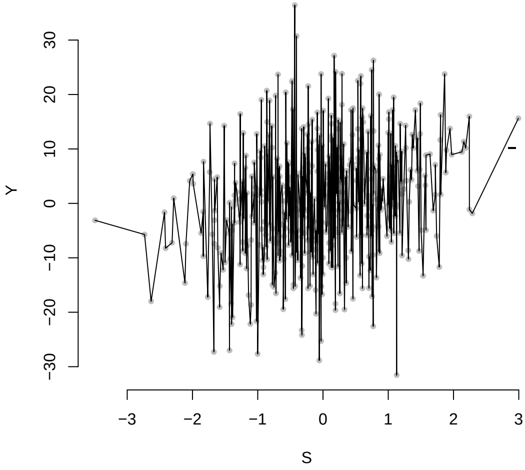
What does this say?

- Stratifying (conditioning): examining means of Y given values of S
- Borrowing information across bins:
  - Assuming that means have a global and linear movement with S, β_1

Linear Regression

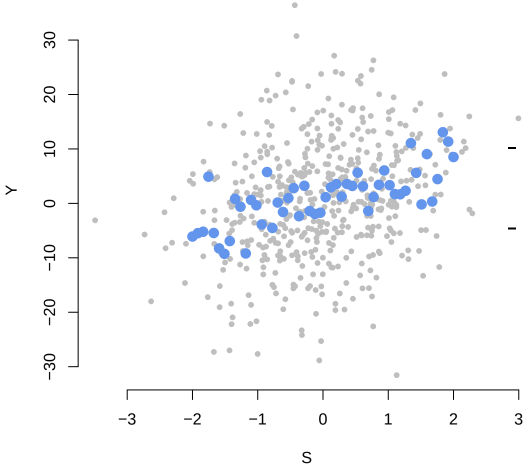


Linear Regression



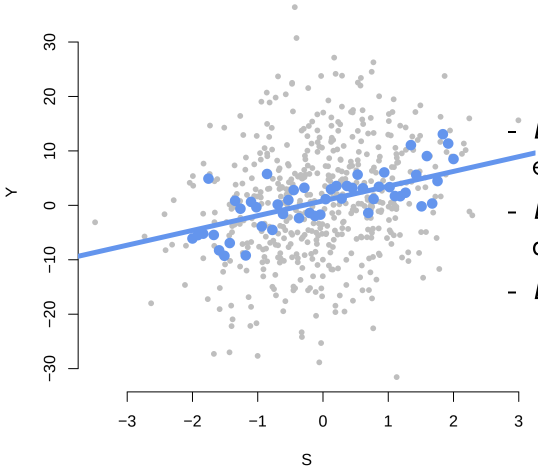
- $E[Y|S]$ with no stratification: every point is in its own "bin"

Linear Regression



- $E[Y|S]$ with no stratification: every point is in its own "bin"
- $E[Y|S]$ after binning the data–no assumed relationship

Linear Regression



- $E[Y|S]$ with no stratification: every point is in its own "bin"
- $E[Y|S]$ after binning the data–no assumed relationship
- $E[Y|S] = \beta_0 + \beta_1 S_i$

For every random variable Y and S we can always write the random variable as,

$$(0.1)$$

For every random variable Y and S we can always write the random variable as,

$$Y_i = \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \quad (0.1)$$

For every random variable Y and S we can always write the random variable as,

$$
\begin{aligned}
Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\
Y_i &= E[Y|S] + \epsilon_i
\end{aligned}
\tag{0.1}
$$

For every random variable Y and S we can always write the random variable as,

$$\begin{aligned} Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\ Y_i &= E[Y|S] + \epsilon_i \\ Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i \end{aligned} \quad (0.1)$$

For every random variable Y and S we can always write the random variable as,

$$
\begin{aligned}
Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\
Y_i &= E[Y|S] + \epsilon_i \\
Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i
\end{aligned}
$$

Where we have used our assumption about $E[Y|S] = \beta_0 + \beta_1 S_i$

(0.1)

For every random variable Y and S we can always write the random variable as,

$$\begin{aligned} Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\ Y_i &= E[Y|S] + \epsilon_i \\ Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i \end{aligned}$$

Where we have used our assumption about $E[Y|S] = \beta_0 + \beta_1 S_i$
We'll define our residuals to be,

(0.1)

For every random variable Y and S we can always write the random variable as,

$$
\begin{aligned}
Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\
Y_i &= E[Y|S] + \epsilon_i \\
Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i
\end{aligned}
$$

Where we have used our assumption about $E[Y|S] = \beta_0 + \beta_1 S_i$
We'll define our residuals to be,

$$
\epsilon_i = Y_i - (E[Y|X]) \qquad (0.1)
$$

For every random variable Y and S we can always write the random variable as,

$$
\begin{aligned}
Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\
Y_i &= E[Y|S] + \epsilon_i \\
Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i
\end{aligned}
$$

Where we have used our assumption about $E[Y|S] = \beta_0 + \beta_1 S_i$
We'll define our residuals to be,

$$
\begin{aligned}
\epsilon_i &= Y_i - (E[Y|X]) \\
&= Y_i - (\beta_0 + \beta_1 S_i)
\end{aligned} \tag{0.1}
$$

For every random variable Y and S we can always write the random variable as,

$$
\begin{aligned}
Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\
Y_i &= E[Y|S] + \epsilon_i \\
Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i
\end{aligned}
$$

Where we have used our assumption about $E[Y|S] = \beta_0 + \beta_1 S_i$
We'll define our residuals to be,

$$
\begin{aligned}
\epsilon_i &= Y_i - (E[Y|X]) \\
&= Y_i - (\beta_0 + \beta_1 S_i)
\end{aligned}
$$

We are going to find the β_0^*, β_1^* that minimize the sum of squared residuals,

(0.1)

For every random variable Y and S we can always write the random variable as,

$$\begin{aligned} Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\ Y_i &= E[Y|S] + \epsilon_i \\ Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i \end{aligned}$$

Where we have used our assumption about $E[Y|S] = \beta_0 + \beta_1 S_i$
We'll define our residuals to be,

$$\begin{aligned} \epsilon_i &= Y_i - (E[Y|X]) \\ &= Y_i - (\beta_0 + \beta_1 S_i) \end{aligned}$$

We are going to find the β_0^*, β_1^* that minimize the sum of squared residuals,

$$(\beta_0^*, \beta_1^*) = \texttt{argmin}_{\beta_0, \beta_1} \sum_{i=1}^{N} \epsilon_i^2 \quad (0.1)$$

For every random variable Y and S we can always write the random variable as,

$$\begin{aligned} Y_i &= \underbrace{Y_i - E[Y|S]}_{\epsilon_i} + E[Y|S] \\ Y_i &= E[Y|S] + \epsilon_i \\ Y_i &= \beta_0 + \beta_1 S_i + \epsilon_i \end{aligned}$$

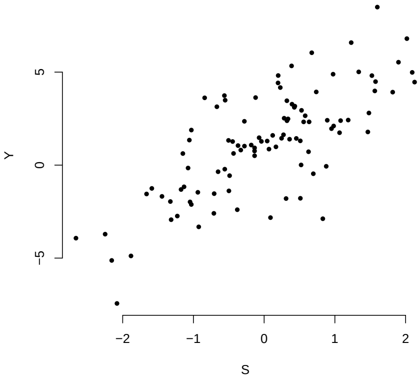
Where we have used our assumption about $E[Y|S] = \beta_0 + \beta_1 S_i$
We'll define our residuals to be,

$$\begin{aligned} \epsilon_i &= Y_i - (E[Y|X]) \\ &= Y_i - (\beta_0 + \beta_1 S_i) \end{aligned}$$

We are going to find the β_0^*, β_1^* that minimize the sum of squared residuals,

$$(\beta_0^*, \beta_1^*) = \texttt{argmin}_{\beta_0, \beta_1} \sum_{i=1}^{N} (Y_i - \beta_0 - \beta_1 S_i)^2$$

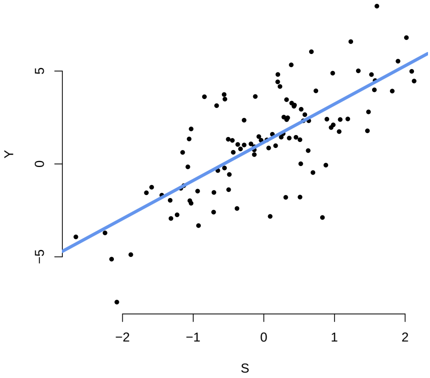
Graphically:



$$\beta_0^* = \overline{Y} - \beta_1^* \overline{S}$$
$$\beta_1^* = \frac{\sum_{i=1}^{N}(S_i - \overline{S})Y_i}{\sum_{i=1}^{n}(S_i - \overline{S})^2}$$
$$\beta_1^* = \frac{cov(S, Y)}{var(S)}$$

```
regress<- lm(Y S)
```

Graphically:



$$\beta_0^* = \overline{Y} - \beta_1^* \overline{S}$$

$$\beta_1^* = \frac{\sum_{i=1}^{N}(S_i - \overline{S})Y_i}{\sum_{i=1}^{n}(S_i - \overline{S})^2}$$

$$\beta_1^* = \frac{cov(S, Y)}{var(S)}$$

```
regress<- lm(Y S)
```

How To Interpret a Regression?

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.

Common to specify,

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.
Common to specify,

$$Y_i = \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.

Common to specify,

$$Y_i = \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.

Common to specify,

$$Y_i = \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE
- α will be a consistent estimate of ATE (converge in probability) if there are no omitted variables

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.
Common to specify,

$$Y_i \quad = \quad \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE
- α will be a consistent estimate of ATE (converge in probability) if there are no omitted variables
- α will be a consistent estimate of ATE (converge in probability) if treatment is as good as randomly assigned, given model

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.
Common to specify,

$$Y_i = \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE
- α will be a consistent estimate of ATE (converge in probability) if there are no omitted variables
- α will be a consistent estimate of ATE (converge in probability) if treatment is as good as randomly assigned, given model
    1) X's are pre-treatment (not consequences of treatment)

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.
Common to specify,

$$Y_i = \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE
- α will be a consistent estimate of ATE (converge in probability) if there are no omitted variables
- α will be a consistent estimate of ATE (converge in probability) if treatment is as good as randomly assigned, given model
  1) X's are pre-treatment (not consequences of treatment)
  2) We know the X's that are related to treatment assignment + outcome

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.
Common to specify,

$$Y_i = \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE
- α will be a consistent estimate of ATE (converge in probability) if there are no omitted variables
- α will be a consistent estimate of ATE (converge in probability) if treatment is as good as randomly assigned, given model
  1) X's are pre-treatment (not consequences of treatment)
  2) We know the X's that are related to treatment assignment + outcome
  3) The linear model is right (not more complex functional forms)

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.
Common to specify,

$$Y_i = \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE
- α will be a consistent estimate of ATE (converge in probability) if there are no omitted variables
- α will be a consistent estimate of ATE (converge in probability) if treatment is as good as randomly assigned, given model
  1) X's are pre-treatment (not consequences of treatment)
  2) We know the X's that are related to treatment assignment + outcome
  3) The linear model is right (not more complex functional forms)

You get one causal effect per regression:

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.
Common to specify,

$$Y_i \quad = \quad \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE
- α will be a consistent estimate of ATE (converge in probability) if there are no omitted variables
- α will be a consistent estimate of ATE (converge in probability) if treatment is as good as randomly assigned, given model
  1) X's are pre-treatment (not consequences of treatment)
  2) We know the X's that are related to treatment assignment + outcome
  3) The linear model is right (not more complex functional forms)

You get one causal effect per regression:

- The whole point of the X's is just to replicate experimental conditions

How To Interpret a Regression?

Suppose (again) dichotomous treatment T_i and a host of covariates $X_{i1}, X_{i2}, \ldots, X_{iK}$.
Common to specify,

$$Y_i = \beta_0 + \alpha T_i + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_K X_k + \epsilon_i$$

- α is an estimate of the ATE
- α will be a consistent estimate of ATE (converge in probability) if there are no omitted variables
- α will be a consistent estimate of ATE (converge in probability) if treatment is as good as randomly assigned, given model
  1) X's are pre-treatment (not consequences of treatment)
  2) We know the X's that are related to treatment assignment + outcome
  3) The linear model is right (not more complex functional forms)

You get one causal effect per regression:

- The whole point of the X's is just to replicate experimental conditions
- Not to estimate separate causal effects

TABLE 4. Panel Regression of County-level Unemployment and the Democratic Percent of the Two-party Vote for President, 1996–2008

	I Coefficient (standard error)[a]	II Coefficient (standard error)	III Coefficient (standard error)	IV Coefficient (standard error)
Unemployed (percent in county)	.176**		.203***	.198***
	(.086)		(.061)	(.058)
Change in unemployment from previous year		.113		.079
		(.105)		(.101)
State unemployment		−.067	−.229	−.243
		(.613)	(.687)	(.589)
Median household income ($1000s)	.030	.015	.030	.028
	(.036)	(.037)	(.037)	(.037)
Democratic vote in previous election	.906***	.914***	.905***	.906***
	(.018)	(.018)	(.084)	(.019)
Percent urban	.017	.017*	.017*	.017*
	(.009)	(.010)	(.010)	(.010)
Percent African American	.096***	.096***	.095***	.095***
	(.009)	(.008)	(.008)	(.008)
Percent without high school diploma	.085***	.097***	.083**	.085**
	(.030)	(.033)	(.035)	(.034)
Percent with four-year college degree or more	.130***	.129***	.132**	.132*
	(.047)	(.045)	(.053)	(.052)
Percent aged 18–30	.029	.013	.029	.026
	(.025)	(.027)	(.030)	(.029)
Percent 65 or older	−.013	−.029	−.013	−.016
	(.021)	(.023)	(.025)	(.024)
Constant	−8.29***	−6.67**	−7.30**	−7.14**
	(1.27)	(3.23)	(3.39)	(3.06)
N	12,444	12,444	12,444	12,444
N counties	3,111	3,111	3,111	3,111
N years (fixed)[b]	4	4	4	4
R-squared: overall (within years)	.84 (.92)	.84 (.91)	.84 (.92)	.84 (.92)

[a]Standard errors are bootstrapped with 250 replications; significance tests based on the normal distribution.
[b]State effects were fixed through inclusion of state-dummy variables not reported here.
*statistical significance at .10, two-tailed test;

See you in the spring!!